The Official Journal of the International Society on Infant Studies Infancy

Infancy, ***16***(1), 39–44, 2011

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/j.1532-7078.2010.00053.x

WILEY-BLACKWELL

Perceptual Organization of Visual Structure Requires a Flexible Learning Mechanism

Richard N. Aslin
Department of Brain and Cognitive Sciences
University of Rochester

Bhatt and Quinn (2011) provide a compelling and comprehensive review of empirical evidence that supports the operation of principles of perceptual organization in young infants. These principles, which are largely synonymous with the Gestalt laws of visual perception, raise a number of important, and as yet unanswered questions. Bhatt and Quinn tackle some of these questions and propose several mechanisms by which infants progress from a few principles of perceptual organization to the full set of principles that operate in adults.

The first question is about origins: Are any of the principles of perceptual organization present at birth in the absence of visual experience? Although the Gestalt psychologists assumed that all the principles of perceptual organization were innate, presumably because of evolutionary pressures to ensure their presence in the species, these early 20th century psychologists did not have the methods required to measure the operation of these principles in young infants. Bhatt and Quinn (2011) summarize evidence that by 3–4 months of age there are at least two such principles of perceptual organization in operation: common motion and good continuation. Unfortunately, these data from 3- to 4-month-olds do not confirm a nativist origin because 3–4 months of postnatal visual experience provides an enormous opportunity for the learning of these principles. Demonstrations of

Correspondence should be sent to Richard N. Aslin, Department of Brain and Cognitive Sciences, Meliora Hall, River Campus, University of Rochester, Rochester, NY 14627. E-mail: aslin@cvs.rochester.edu

rapid statistical learning (e.g., Saffran, Aslin, & Newport, 1996) provide a cautionary note to those who believe infants cannot plausibly acquire complex information from experience in the first few postnatal months.

Given that some principles of perceptual organization are clearly acquired later in development, the second part of the origins question is whether the emergence of the full complement of organizational principles comes about from maturation, independent of visual experience, or whether experience is necessary for the progression from two principles to all the principles of perceptual organization? Unfortunately, we cannot easily determine the answer to this second part of the origins question because it would entail depriving infants of certain types of visual experience. But we do know, as summarized by Bhatt and Quinn (2011), that infants *can* learn some principles of perceptual organization in a laboratory setting (c.f., Quinn & Bhatt, 2005). These laboratory demonstrations do not force the conclusion that similar learning processes occur in the natural environment, but they are consistent with such a conclusion and clearly demonstrate the *capacity* of such a learning mechanism.

The second question addressed by Bhatt and Quinn (2011) is how the principles of perceptual organization—most of the principles if some are innate or all of them if none are innate—are learned. They propose five *types* of learning experiences that are relevant to acquiring the principles of perceptual organization: (1) exposure to variability, (2) scaffolding from an already functioning principle to a related principle, (3) exposure to holistic images, (4) exposure to features, and (5) extraction of feature correlations. Bhatt and Quinn also propose two domain-general *mechanisms* that mediate learning: (1) attentional enhancement, and (2) unitization.

It is undeniable that these types of learning experiences, and their proposed underlying mechanisms, *could* play a role in the developmental acquisition of some or all of the principles of perceptual organization. However, *seven* such experiences/mechanisms is quite a large number, and they have the flavor of an exhaustive list rather than an integrated model. Might there be simplifications in terminology and efficiencies in computational principles that could account just as well for the phenomena described by Bhatt and Quinn (2011)? I think there are, and they fall under the rubric of Bayesian approaches to learning.

The essence of the Bayesian perspective (see Gopnik et al., 2004) is that learners are attempting, either implicitly or explicitly, to make predictions about the world based on two sources of information. One source consists of the observables—called likelihoods—and the other source consists of expectations—call prior probabilities. Predictions refer to the assignment of a probability that an observed event has a specific cause. Of course, you cannot observe the cause itself, so an inference is made about the cause based on the likelihood and the prior. Specifically, the prediction—called the posterior probability—is the product of the likelihood and the prior. Since the likelihood can be observed under some conditions (e.g., when you know the cause and observe the effect), and the prior is simply your bias about the cause based on previous experience, both of these terms in the equation (likelihood × prior) can be measured and used to estimate the posterior probability.

Given this very brief summary of Bayes Law (posterior probability = likelihood × prior), let's return to Bhatt and Quinn's (2011) seven experiences/mechanisms. *Exposure to variability* and *attentional enhancement* refer to a fundamental tenet of Bayesian learning: When combining information from two or more sources, weight each information source in proportion to its reliability (or the inverse of its variability). Thus, if a cue is highly variable (i.e., unreliable), all other things being equal, down-weight it and up-weight all other cues.

Exposure to holistic images (i.e., whole objects), *exposure to features* (i.e., parts of objects), and *extraction of feature correlations* also refer to the foregoing Bayesian process of cue re-weighting. If in the context of a set of exemplars, there is greater reliability (less variability) at the level of wholes than at the level of parts (or vice versa), then an ideal learner should up-weight the information contained at that level of the visual information hierarchy. And if a higher-order statistic better describes the structural relations present in the set of exemplars than a lower-order statistic, then an ideal learner should rely more on that level of the visual information hierarchy. Moreover, the mechanism of *unitization* is simply an extension of the process of extracting feature correlations by building up larger chunks from statistically coherent subgroups of elements (c.f., Fiser & Aslin, 2002; Orban, Fiser, Aslin, & Lengyel, 2008). But again, whether the learner weights the parts more than the wholes depends on how reliable those two levels of information are for a given set of features/objects and for a given ability to detect and discriminate among those features/objects. For example, if the features were blurred, or the learner could not resolve their fine detail, then these features would be unreliable and would likely be dominated by more global information.

The final type of learning experience—*scaffolding* from an already functioning principle to a related principle—is perhaps the most powerful and least understood of those discussed by Bhatt and Quinn (2011). How is it that this scaffolding process, or bootstrapping, enables the learner to transfer knowledge from one context to another? In my judgment, the key to any robust learning mechanism is flexibility: The information that is learned should not be so narrow as to disallow the process of generalization, yet not so broad that it applies to everything. In Bayesian terminology, this is called the automatic Occam's razor principle—the learner's model of the world is broad enough to generalize to novel exemplars, but narrow enough to prevent over-generalization to domains that are better fit by having two models rather than one (MacKay, 1992).

Having two models rather than one is directly relevant to the studies of grouping principles reviewed by Bhatt and Quinn (2011). Why would adults perceive a circular configuration of square-shaped elements as a circle? Presumably, the adults are basing their judgment on the global shape and not the local (element) shape. Of course, adults can be instructed to attend to either the global or the local shape, but infants cannot be so instructed. Why then would infants attend more reliably to global shape? Again, the answer may reside in how infants resolve the competing alternative interpretations based on the reliability of cues in their visual world. As noted earlier, global shape is more robust to noise or blur—that is, it is a more reliable source of information—than local shape.

But now consider a display of alternating vertical columns of Xs and Os. Why would infants perceive this display as more similar to a vertical set of black and white bars than to a horizontal set of black and white bars? Adults presumably perform this similarity judgment based on extracting the similarity of the vertically adjacent element-shapes and generalizing to the vertical black and white bars. Infants do not show such generalization naturally until 6–7 months of age (Quinn, Bhatt, Brush, Grimes, & Sharpnack, 2002). And as noted by Bhatt and Quinn (2011), this failure to generalize is not the result of an inability to discriminate the Xs from the Os.

From the Bayesian perspective, both of the foregoing examples—the circle of square-elements and the generalization from columns of Xs and Os to alternating vertical stripes—are cue-competition tasks. There are two or more sources of information that have incompatible underlying causes. In the case of the circle of squares, infants can attend to the local shapes or the global configuration. And in the case of the columns of Xs and Os, infants can attend to the vertical array of elements (if they also attend to shape similarity in each column) or to the horizontal array of elements (if they ignore shape similarity).

How do infants resolve a cue-competition situation? As noted earlier, cues are weighted in proportion to their reliability, and when there are two or more conflicting cues, they are combined by performing a weighted average. Thus, if one cue is quite weak and a second cue is quite strong, the combined outcome will favor the stronger cue (see Jacobs, 2002). What, then, enables older infants to learn that columns of Xs and Os should generalize to vertical stripes? According to the Bayesian perspective, there are two answers to this question. First, the learner's model of the world (in this case, the world consists of rows and columns)—which is captured by the predictions entailed in the posterior probabilities—is based on the product of the likelihood and the priors. Each time new information about the causal nature of the world is acquired from experience, the prior is updated because that prediction is now more expected even in the absence of any further data. This process is called iterative Bayes (see Kruschke, 2006) because learning occurs via cycles of (data × priors) followed by updated priors and then the next sample of (data × priors), etc. Notably, the initial state of the priors represents innate biases before there has been any learning experience.

Second, the learner's model of the world can be hierarchical (see Tenenbaum, Griffiths, & Kemp, 2006). For example, the circle of squares is *both* circular and square at different levels of the visual feature hierarchy. Thus, one way to handle this Bayesian hierarchy is to entertain (in an implicit sense) two models of the world and allow the best model to control the outcome of perception for a given task. An interesting corollary of these competing models is that they may share *some* common causes. If so, then the models not only compete at the level of differences, but they also generalize at the level of these commonalities. Thus, a class of Bayesian models can learn to learn (see Kemp, Goodman, & Tenenbaum, 2010).

In sum, Bhatt and Quinn (2011) have done an excellent job of laying out the empirical data on infants' use of principles of perceptual organization in the visual domain. They also have provided a comprehensive list of experiences that could serve to trigger the learning of at least some of these principles of perceptual organization, and two domain-general mechanisms that could operate in the learning process itself. What I have offered in this commentary is the prospect that the computational principles of a Bayesian approach to learning could provide a unifying and simplified account of the seven experiences/mechanisms offered by Bhatt and Quinn. It remains to be seen by future empirical work and computational modeling whether these Bayesian principles provide a quantitative fit with the empirical data and a plausible implementation at the level of learning mechanisms.

REFERENCES

Bhatt, R. S., & Quinn, P. C. (2011). How does learning impact development in infancy? The case of perceptual organization. *Infancy*, *16*, 2–38.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*, 15822–15826.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1–31.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*, 345–350.

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, *34*, 1185–1243.

Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review*, *113*(4), 677–699.

MacKay, D. J. C. (1992). Bayesian interpolation. *Neural Computation*, *4*, 415–447.

Orban, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*, 2745–2750.

Quinn, P. C., & Bhatt, R. S. (2005). Learning perceptual organization in infancy. *Psychological Science*, *16*, 515–519.

Quinn, P. C., Bhatt, R. S., Brush, D., Grimes, A., & Sharpnack, H. (2002). Development of form similarity as a Gestalt grouping principle in infancy. *Psychological Science*, *13*, 320–328.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Science*, *10*, 309–318.